QA
76
M94
SI-Ref.

M U D P I E no. 3

Museum and University Data, Program, and Information Exchange

SYMPOSIUM ON INFORMATION PROBLEMS IN NATURAL SCIENCES

This symposium was held December 18-20, in Mexico City, under the sponsorship of the Universidad Nacional de Mexico and the Smithsonian Institution. It brought together about 60 people with mutual interest in methods of keeping up with the large amounts of information available and accumulating. The focus on various methods of coping with data pertaining to large collections, and work in progress was discussed. The storage of information by museums, such as catalogue records, and the rapid retrieval of this information on demand, received considerable attention. Several papers covered attempts to make bibliographic materials more easy to manipulate. Several methods of preparing, handling, and writing systematic keys were described, as was work with storage and retrieval of data derived from specific research projects. Conclusions drawn by one participant [J. A. Peters] after three days of papers and discussion include the following (it should be clear that these conclusions probably reflect my biases, and cannot be interpreted as representing the results of the symposium in any way):

1. There are two distinct approaches to storage and retrieval of data: the first involves information pertinent to a specific problem, but of such a magnitude that it is difficult if not impossible to handle it without electronic methods; the second involves information that is not immediately in demand, and not necessarily pertinent to any existing problem, but that represents the total body of information about a specific group of objects, with constant demand for changing parts of that body, although its total is never required at any one time. The first approach is used by an individual or group in a specific research project, and the stored information, finally summarized in a single or series of publications, might as well be discarded upon completion. The second is used by many different workers interested in a multitude of projects, is constantly growing as a consequence of additional collections of objects, and its value increases both with use and with age.

2. It is now completely feasible to perform both tasks, in the sense that computers have the capacity to handle the masses of information. Use of computer methods seem to be equally practical and wise, in fact almost mandatory, in the first approach mentioned above. There was, however, a considerable division of opinion concerning the practicality and wisdom of computer use in the second approach. It seems clear that it can be practical, in the sense that money has been made available for several such projects currently in progress, although it has so far invariably been funded by sources extraneous to the budget of the museum or university concerned. Representatives of smaller museums repeatedly pointed out that it cannot be practical for their institutions, because their budgets will never be large enough. Even were one to assume complete

SMITHSONIAN
1968
LIBRARIES

practicality of such automation, however, the challenges as to its wisdom persist. These challenges include problems of reliability of taxonomic determinations, accuracy of data in the catalogues, frequency of demand by users, the non-repetitive nature of the data, and others. It is clear that these challenges will not be answered until the supporters of the second approach have supplied acceptable answers to some of the day-to-day queries concerning the collections of a museum, and have given the scientific community accurate, detailed, and non-ambiguous data on costs.

3. It is apparent that a considerable amount of duplication of effort is taking place in the area of storage and retrieval of museum data. Many different institutions or individual workers are preparing methods for encoding locality data, scientific names, bibliographic information, and in many cases the methods will not be compatible with each other in the future. Worse still, errors are made again and again that have been experienced and solved by earlier workers, but require repeated laborious solutions by the next generation. Close coordination of work on mutual problems is mandatory.

In the hopes that the abstracts of the program might prove of interest and perhaps of use to the readers of MUDPIE, they are reproduced in part on the following pages. The rest will be included with a later issue.

Smithsonian Institution
January 10, 1968

2. "AN INFORMATION STORAGE AND RETRIEVAL SYSTEM FOR BIOLOGICAL AND GEOLOGICAL DATA-DESIGN CONSIDERATIONS"

Reginald Creighton
Smithsonian Institution
Washington, D.C.

The system design attempts to achieve optimum processing economy while permitting the system user maximum freedom of activity regarding data preparation, query expression and output presentation. Due to the disparity of data parameters involved (birds, crustecia, and rocks) a completely generalized software system is untenable. A technique for retrieval program generation was devised wherein a custom program, tailored to the specific queries submitted, and to the specific data format involved, is created on each processing cycle.

Despite the diversity of data parameters encompassed, four major elements of similarity have been identified. All participants wish to access data by taxa name, by geographical or political location; by catalog number; and in certain cases by stratigraphic designators. The data base comprised of specimen and bibliographic data is maintained in a dynamic phylogenetic sequence and is accessed by four indices: phylogenetic, geopolitical, catalog number (intervals) and stratigraphic. Other basic tenets of the system include:

1. All consistently formated data is searchable.
2. No abbreviation or user coding is imposed by the system. The computer translates natural language input into code and retranslates output.
3. User training is minimized.
4. A machine independent language is used (to enable exchange of software as well as of data).

The system is now being improved and implemented by the Information Systems Division, Smithsonian Institution, Washington, D.C.

KEYWORD ELEMENTS:
Retrieval program generation,dynamic phylogeny,human factors of query,machine independent query,generalized data collection,locational heirarchy,common denominator indices,synonomy rectification,heirarchical indices, biological numericlature, query batch optimization.

3. "PREPARACION Y MANIPULACION DE CLAVES SISTEMATICAS UTILIZANDO COMPUTADORAS DE TIEMPO COMPARTIDO"

James A. Peters
Smithsonian Institution
Washington, D.C.

Se ha construído una clave para identificación de los géneros de culebras neotrópicas utilizando una computadora de tiempo compartido y el lenguaje llamado "BASIC". La simplicidad de este lenguaje ha permitido a la computadora usar solamente catorce características, las cuales proveen toda la información necesaria para las identificaciones. Se pueden incluír otras características dentro del programa cuando hay un género único, y la computadora escribe la información adicional en inglés, con los nombres genéricos pertinentes. Cuando hay un género nuevo o uno no incluído anteriormente en la clave, se puede incluir con gran facilidad. Necesita nada más que desafiar a la computadora con las catorce características del nuevo género, porque ésta aprovecha un nombre genérico antiguo. Se puede comparar ésta con el género nuevo, y distinguir entre los dos para facilitar la introducción del nuevo. Todos los géneros conocidos hasta el momento de Guatemala hasta Argentina están en la clave. Otros especialistas que quieran usar la clave, pueden pedir directamente una copia que sea aceptada por la computadora, o, si no tienen una computadora de tiempo compartido, ellos pueden mandar una lista de los ejemplares con las catorce características para desafiar a nuestra computadora.

4. "A COMPUTERIZED INFORMATION RETRIEVAL SYSTEM FOR TAXONOMY"

D. J. Rogers
University of Colorado
Boulder, Colorado

5. "COMPUTER APPLICATIONS IN TAXONOMIC LITERATURE"

Gilbert S. Daniels
Hunt Botanical Library
Pittsburgh, Pa.

Bibliographical research at the Hunt Botanical Library on the early literature of botany has been organized to make maximum use of computer data handling. Lengthy citations for more than 25,000 volumes are stored on computer tapes and are updated and recollated as further research is carried out on each volume, Data on additional volumes can be merged as it becomes available. Sorting can be done on any category of information contained in the individual volume records with hard copy produced by the computer in a form suitable for direct publication by photo offset. Ancillary research projects not conceived of initially are greatly facilitated by the versatile access and sorting provided by the computer.

Further applications and implications to taxonomic literature studies are discussed in this paper.

6. "SOME ASPECTS OF ON LINE INFORMATION RETRIEVAL LANGUAGES"

G.K. Hutchinson
Computer Center
Texas Technological College
Lubbock, Texas.

Information retrieval techniques of interest to museums include: 1) KWIC, 2) Current Awareness, 3) Qwick Qwery, and 4) On Line Dialogues. The on-line search is potentially the most powerful, but requires considerable effort and thoughtful consideration as to its language design, hardware configuration and, perhaps more important, the matching of the language to the hardware via the software implementation. Retrieval techniques will be briefly reviewed. Implementation considerations and evaluation criteria will be discussed.

7. "TELECOMMUNICATION AND ON LINE ACCESS TO COMPUTERS"

Nicholas J. Suszynski
Smithsonian Institution
Washington, D.C.

Direct access to the computer from the deskside of a user who has a need to process his data on a computer is highly desirable. Several manufacturers make remote terminals for the computers which facilitate such interaction. Video data terminals (basically a television tube with a typewriter keyboard) connected to a computer by way of a telephone line is by far the best method devised as yet for inputting data to the computer and for making requests against the computer based data bank. Video data terminals are also considerably faster and generally more accurate than other means of inputting data (keypunches, verifiers etc.). If this is the case, why then aren't they used more widely?.

There are several reasons for scarcity of remote usage and the two major ones are discussed in depth. Despite the pressure from the computer industry and the users, the common carrier monopoly is slow in improving its telecommunication capability. The common carriers 10 + 10 cycle (ten years in development followed by ten years of manufacturing as compared to computer industry of 5 + 5) will continue into the forseeable future. This means that current characteristics of the communication channel will change very slowly (noise level, frequency response, capacity of channel etc.).

Also, since the users of computer data banks will be separated from their computers by tens and hundreds and often thousands of miles, economic impact of the telecommunication costs is crucial. The technology of data transmission is discussed only briefly and main concentration is on economic aspects of it, showing with slides and charts costs of various lines in relation to the distance.

The second major obstacle to remote multi-processing is the scarcity of the computer systems (hardware and software) capable of economic processing in the foreground of remote messages, and in the backgground of batch processing the normal load of the data center (such things as updating of the files with most current information etc.). There are isolated exceptions to this, and they are in the area of dedicated systems; systems that are offered for a special mission such as airline reservation systems, or a demand deposit accounting system in the banking industry, or general purpose systems having few options with many limitations which normally are offered in a large city in order to capitalize on the relatively low cost of the telephone usage. These general purpose systems are usually offered by the service bureaus.

The paper will strive to show that nation-wide computer utility with direct access to various data banks, although technologically feasible, is not economical as yet. That the biggest single cost, in addition to the initial data conversion needed to establish the data bank, is the cost of data transmission. That because of it, these networks will form at first in the large cities where the cost of a local telephone call is negligible. In the immediate future (2 to 3 years), rather advanced and economical time-shared systems will be available within cities. Their use outside the city will depend largely on the size of the telephone bill that will come with it. An overview of currently available commercial systems will be presented.

8. "THE OKLAHOMA MUSEUMS INVENTORY PROJECT"

A. F. Ricciardelli
University of Oklahoma
Norman, Oklahoma

During the period 1965-67, a pilot study utilizing the museums in the State of Oklahoma was conducted to devise a workable system for inventorying ethnological collections. It is hoped that such a system may be applicable on a nation wide basis and a union index of ethnological materials created. Specifically, this study investigated such matters as the time and cost required to inventory ethnological specimens to determine whether an inventory procedure could be devised which would insure reliability and consistency in the data collected, and to determine the most efficient means for storage and retrieval of that data, including automatic data processing systems. Extensive data has been collected on these topics and various conclusions and recommendations are presented. As a by-product of this study, a workable file containing the data for the Oklahoma museums now exists at the Stovall Museum of the University of Oklahoma and is available for use by scholars.

9. "A GENERALIZED APPROACH TO INFORMATION RETRIEVAL: THE BIRD AND TAXIR SYSTEMS"

Robert Brill
University of Colorado
Boulder, Colorado

10. "A WORLD PLANT GERM PLASM RECORD SYSTEM"

C. F. Konsak, Washington State University
K. W. Finlay, University of Adelaide, Australia
B. Sigurbjörnsson, International Atomic Energy Agency, Vienna, Austria and
G. Delhove, FAO, Roma Italia.

Studies are now well advenced on an internationally coordinated system adapted for the storege, retrieval end processing of records end data by computers. These studies are being conducted by a working group on international standardization in crop research data recording. This working group was established on the recommendation of a group of scientists assembled at IAEA headquarters in Vienna in December, 1965, at the invitation of the Directors General of FAO and IAEA. The working group was given the task of investigating the possibilities for standardizetion of data recording procedures and developing a model system for records on accessions of food, feed and fiber plants.

A series of studies on a preliminary trial system for storage and retrieval of records by computer were successfully completed in february, 1967 by the coordinator and colleagues with the guidance and assistance of other working group members and the IAEA computer staff. A number of preliminary recommendations on data recording procedures were prepared for review. The FAO-IAEA Working Group on Standardization in Crop Research Data Recordingwill also participete in the International Biological Program (IBP) Section on Germ Plasm Exploration, Maintenance and Evaluation. As a part of this program, some standardized procedures have been proposed for records on new germ plasm obtained from plant explorations, from induced mutations and from hybridization or selection programs as well as for records on existing collections of germ plasm being maintained at stations over the world.

A model for the organization and integration of internationally standardized records has been developed and it is proposed to establish a coordinated world plant germ plasm record system, including the central record file at FAO in Rome. Wide scale field tests of the proposed system are being planned. Holders of a number of the major world collections of wheat germ plasm stocks will be asked to participate in this test. The system will also be used in plant exploration and adaptation projects of IBP and in programs coordinated by FAO and FAO/IAEA.

11. "PROYECTO PILOTO DE RECUPERACION AUTOMATICA DE INFORMACIO DEL HERBARIO NACIONAL DE LA U.N.A.M.

L. Alonso, L.Scheinvar y A. Gómez-Pompa, Instituto de Biología
U. N. A. M., México

Se presenta el proyecto piloto de recuperación automática de la información, realizado en el Herbario Nacional del Instituto de Biología.

Se tomaron como base los 3,500 ejemplares de Pteridophyta. Se describe el formato y codificación seguidos con ejemplifica - ción de los sistemas que fueron utilizados.

El programa fue elaborado en lenguaje SPAR y se usó la com - putadora Bendix C D C G 20 del Centro de Cálculo Electrónico de la U. N. A. M.

Se presenta la codificación de las preguntas y algunas res- puestas seleccionadas.

Se mencionan algunas aplicaciones prácticas para el funciona miento de un herbario y su mejor utilización.

Papers 12 and 13 were not abstracted nor presented.